Supplementary Information

Worobey *et al.* "Direct Evidence of Extensive Diversity of HIV-1 in Kinshasa by 1960"

This file contains

Supplementary text

Supplementary Table S1

Supplementary Figs. S1, S2, and S3 and figure legends

Authenticity of the data. Several lines of evidence argue strongly that the sequences recovered from the 1960 sample are genuine and not the result of contamination:

(i) The work was performed in laboratories dedicated to working with ancient/degraded samples, under strict controls for contamination, including physical separation of pre- and post-PCR work and numerous extraction and PCR blanks that were consistently negative;
(ii) The results were reproducible across multiple extractions followed by RT-PCR, cloning, and sequencing;
(iii) The results were independently replicated in SMW's laboratory (Northwestern University) using three blinded sub-samples of lymph node tissue; the correct sample was identified as the positive by RT-PCR and verified by cloning and sequencing (**Table S1**);
(iv) 12 short RNA fragments, covering the *gag*, *pol*, and *env* genes, were successfully amplified by RT-PCR. No HIV-1 amplicons larger than 126 nucleotides (including the longer sets used by Zhu *et al.* to amplify ZR59) were recoverable from the sample, as predicted for a Bouin's-fixed paraffin-embedded specimen;

(v) Similarly, the qPCR assay using a human *B2M* RNA marker confirmed that amplifiable RNA was present in the sample but that its quality was relatively low, as expected for paraffin-embedded tissues;

(vi) As observed with ZR59, the relatively short branch length of the HIV-1 sequences from 1960 is consistent with it being an early strain;

(vii) The sample originated from a patient (an adult with apparent lymph node abnormalities) and a tissue type (lymph node) where HIV-1 RNA could plausibly be present and detectable; and

(viii) All of the paraffin-embedded tissues yielded sequences with reasonable topological and branch length properties (*i.e.*, subtype B sequences from the Canadian sample, and subtype D and A sequences from the three linked to the Democratic Republic of the Congo, each closely related to previously published sequences of Congolese origin).

Table S1. Primers and sequence-confirmed amplification results.

Primer	Primer Sequence 5'-3'	Frag-ment #	DRC 60	BE81	BE85	CAN97
HIVG1 (F)[a]	ACCCACCTATCCCAGTAGGAGAAAT	1	(+)[c]	(+)	(+)	(+)
HIVG2 (R)	GGTCCTTGTCTTATGTCCAGAATGC					
Pol3290F[b]	GCCAGAAAAAGACAGCTGGACTGTCAA	2	(+)[c]	(+)	(+)	(+)
Pol3415R	AGAGCTTTGGYTCCCCTAAGG					
env6371F	CACCACTCTATTTTGTGCATCAG	3	(+)	(+)	(+)	(-)
env6442R	GCATGTGTAGCCCAGACATTAT					
env6445F	GTGTACCCACAGACCCCAAC	4	(+)	(+)	(+)	(-)
env6542R	CTCATGCATTTGTTCTACCATGT					
env6828F	ACACAGGCTTGTCCAAAGGT	5	(+)[c]	(-)	(-)	(-)
env6890R	ACCAGCTGGGGCACAATAAT					
env7468F	CACTCCCATGCAGAATAAAACA	6	(+)[c]	(+)	(+)	(-)
env7535R	AGGGGCATACATTGCTTGTC					
env7717F[d]	CCACCAAGGCAAAGAGAAGA	7	(-)	(+)	(-)	(+)
env7796R	TCCCAAGAACCCAAGGAAC					
env7899F	ATAGAGGCGCAACAGCATCT	8	(-)	(-)	(-)	(-)
env7977R	TTTCCACAGCCAGGACTCTT					
env8047F	TGCCCTGGAACTCTAGTTGG	9	(+)[c]	(+)	(+)	(-)
env8112R	CCATCCAGGTCATGTTCTCC					
env8423F	CGAAGAAGAAGGTGGAGAGC	10	(+)[c]	(+)	(-)	(-)
env8498R	GTCCCAGGCAAGTGCTAAGA					
env7717F[d,e]	CCACCAAGGCAAAGAGAAGA	11	(+)	N/D[f]	N/D	N/D
env7805R	TCCTGCTGCTCCTAAGAACC					
env7771F[d,e]	GAGCTGTCTTCCTTGGGTTCT	12	(+)	N/D	N/D	N/D
env7846R	GCCTGTACCGTCAGCGTTA					
env7835F[d,e]	GACGGTACAGGCCAGACAAT	13	(+)	N/D	N/D	N/D
env7937R	CCAGACCGTGAGTTTCAACA					
env7890F[d,e]	CTGAGGGCTATAGAGGCTCAAC	14	(+)[c]	N/D	N/D	N/D
env7960R	CTTGCCTGGAGCTGTTTAATG					

[a]Previously published [Boni, J. & Schupbach, J. *J. Virol. Methods* **42,** 309 (1993)].
[b]The remaining primer names correspond to the primer's 5' base location in the HXB2 reference strain.
[c]Independently replicated in SMW's laboratory, Northwestern University.
[d]Overlaps ZR59 sequence.
[e]Designed specifically for HIV-1 group M subtype A (after DRC60 was determined to be A-like).
[f]N/D, no data: amplification not attempted.

Fig. S1. The 50% majority-rule consensus tree summarizing the Bayesian MCMC phylogenetic analysis conducted using MrBayes v3.1.2. Branch lengths (substitutions/site) represent the average for that branch in tree sample. The posterior probability for each node is indicated. DRC60 (red), ZR59 (black), and the three control sequences recovered from paraffin-embedded specimens (gray) are depicted with bold branches. DRC60 and the two positive controls from the DRC form monophyletic clades with other DRC sequences, while the Canadian control falls as expected within subtype B. Each control sequence has a reasonably long terminal branch length consistent with its date of sampling. The subtype A/A1 ancestral node used for the regression line dating analysis is indicated. The country of sampling and additional details about the reference strains used can be found at http://hiv.lanl.gov.

Fig. S2. The maximum clade credibility tree from the relaxed clock Bayesian MCMC analysis conducted using BEAST v1.4.7. This is the same tree as shown in **Fig. 2**, but in rectangular form and with sequence names indicated. The country of sampling and additional details about the reference strains used can be found at http://hiv.lanl.gov.

Fig. S3. Bayesian skyline plot of HIV-1 group M. The plot begins at the median posterior TMRCA (1908). The bold line traces the inferred median effective population size over time with the 95% HPD shaded in blue.

Fig. S1

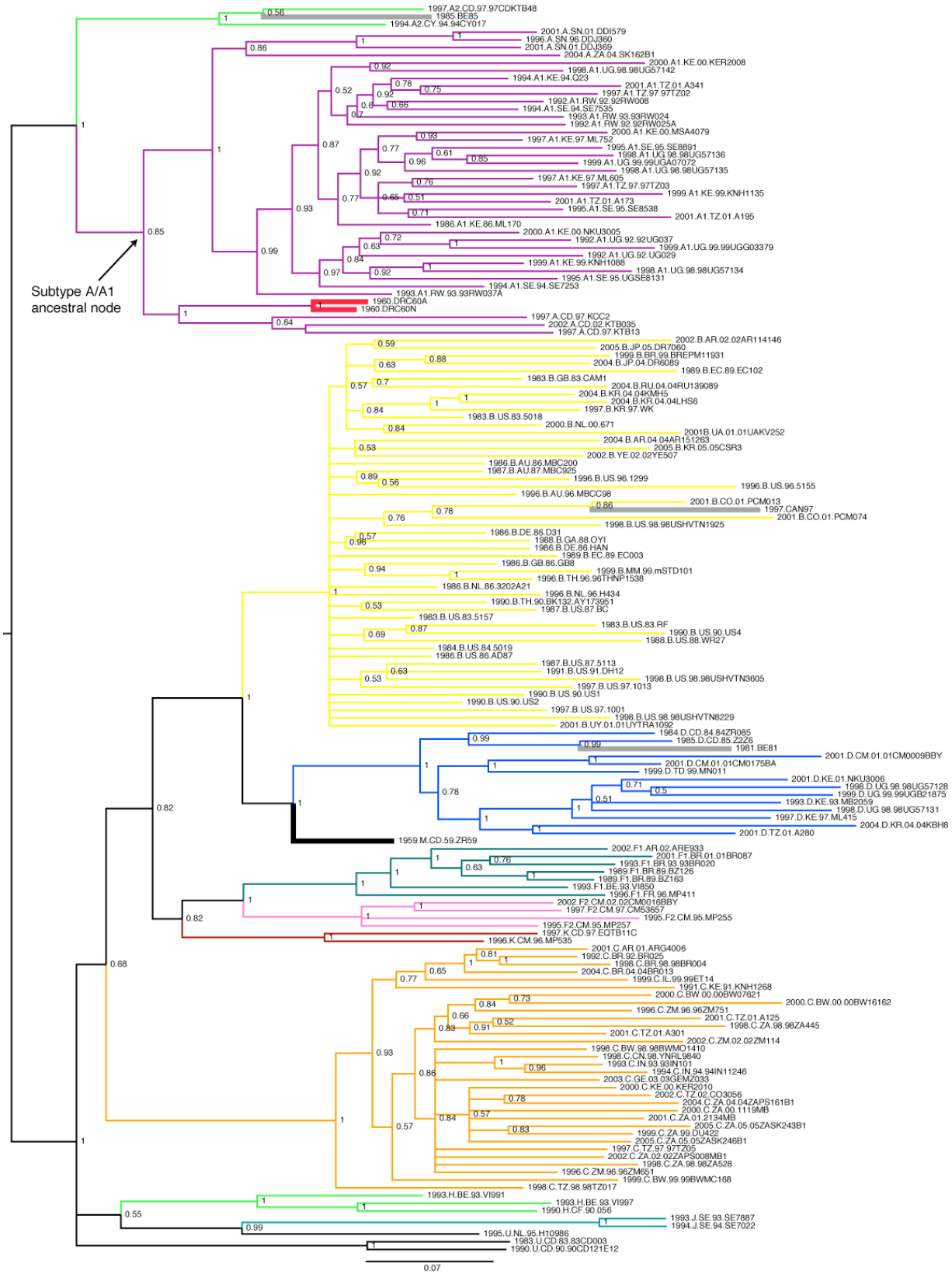
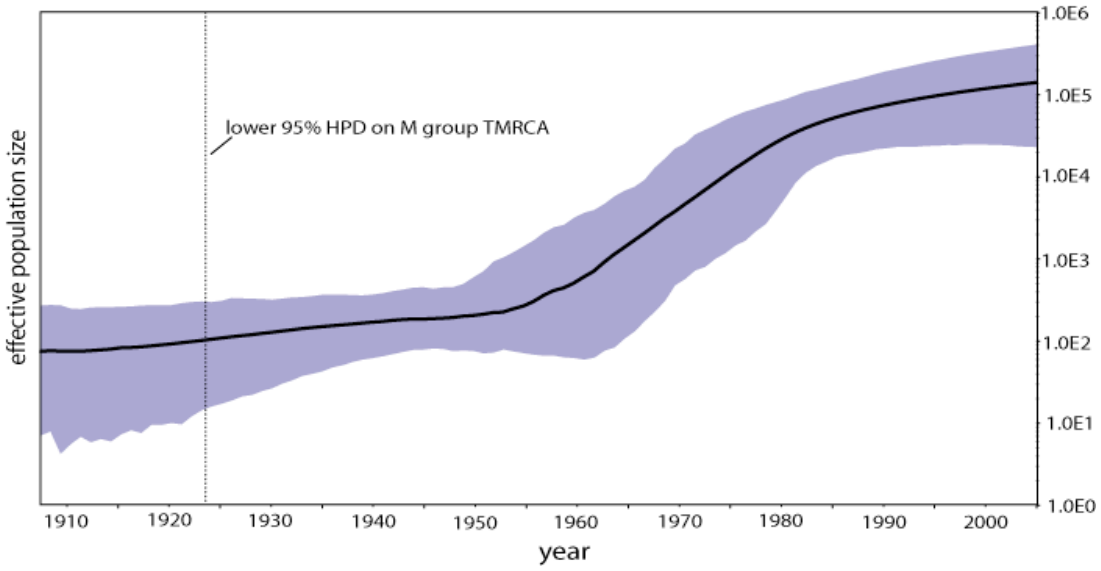

Subtype A/A1
ancestral node
1997.A2.CD.97.97CDKTB48
1985.BE85
1994.A2.CY.94.94CY017
2001.A.SN.01.DDI579
1996.A.SN.96.DDJ360
2001.A.SN.01.DDJ369
2004.A.ZA.04.SK162B1
2000.A1.KE.00.KER2008
1998.A1.UG.98.98UG57142
1994.A1.KE.94.Q23
2001.A1.TZ.01.A341
1997.A1.TZ.97.97TZ02
1992.A1.RW.92.92RW008
1994.A1.SE.94.SE7535
1993.A1.RW.93.93RW024
1992.A1.RW.92.92RW025A
2000.A1.KE.00.MSA4079
1997.A1.KE.97.ML752
1995.A1.SE.95.SE8891
1998.A1.UG.98.98UG57136
1999.A1.UG.99.99UGA07072
1998.A1.UG.98.98UG57135
1997.A1.KE.97.ML605
1997.A1.TZ.97.97TZ03
1999.A1.KE.99.KNH1135
2001.A1.TZ.01.A173
1995.A1.SE.95.SE8538
2001.A1.TZ.01.A195
1986.A1.KE.86.ML170
2000.A1.KE.00.NKU3005
1992.A1.UG.92.92UG037
1999.A1.UG.99.99UGG03379
1992.A1.UG.92.UG029
1999.A1.KE.99.KNH1088
1998.A1.UG.98.98UG57134
1995.A1.SE.95.UGSE8131
1994.A1.SE.94.SE7253
1993.A1.RW.93.93RW037A
1960.DRC60A
1960.DRC60N
1997.A.CD.97.KCC2
2002.A.CD.02.KTB035
1997.A.CD.97.KTB13
2002.B.AR.02.02AR114146
2005.B.JP.05.DR7060
1999.B.BR.99.BREPM11931
2004.B.JP.04.DR6089
1989.B.EC.89.EC102
1983.B.GB.83.CAM1
2004.B.RU.04.04RU139089
2004.B.KR.04.04KMH5
2004.B.KR.04.04LHS6
1997.B.KR.97.WK
1983.B.US.83.5018
2000.B.NL.00.671
2001B.UA.01.01UAKV252
2004.B.AR.04.04AR151263
2005.B.KR.05.05CSR3
2002.B.YE.02.02YE607
1986.B.AU.86.MBC200
1987.B.AU.87.MBC925
1996.B.US.96.1299
1996.B.US.96.5155
1996.B.AU.96.MBCC98
2001.B.CO.01.PCM013
1997.CAN97
2001.B.CO.01.PCM074
1998.B.US.98.98USHVTN1925
1986.B.DE.86.D31
1988.B.GA.88.OYI
1986.B.DE.86.HAN
1989.B.EC.89.EC003
1986.B.GB.86.GB8
1999.B.MM.99.mSTD101
1996.B.TH.96.96THNP1538
1986.B.NL.86.3202A21
1996.B.NL.96.H434
1990.B.TH.90.BK132.AY173951
1987.B.US.87.BC
1983.B.US.83.5157
1983.B.US.83.RF
1990.B.US.90.US4
1988.B.US.88.WR27
1984.B.US.84.5019
1986.B.US.86.AD87
1987.B.US.87.5113
1991.B.US.91.DH12
1998.B.US.98.98USHVTN3605
1997.B.US.97.1013
1990.B.US.90.US1
1990.B.US.90.US2
1997.B.US.97.1001
1998.B.US.98.98USHVTN8229
2001.B.UY.01.01UYTRA1092
1984.D.CD.84.84ZR085
1985.D.CD.85.ZZ26
1981.BE81
2001.D.CM.01.01CM0009BBY
2001.D.CM.01.01CM0175BA
1999.D.TD.99.MN011
2001.D.KE.01.NKU3006
1998.D.UG.98.98UG57128
1999.D.UG.99.99UGB21875
1993.D.KE.93.MB2059
1998.D.UG.98.98UG57131
1997.D.KE.97.ML415
2004.D.KR.04.04KBH8
2001.D.TZ.01.A280
1959.M.CD.59.ZR59
2002.F1.AR.02.ARF933
2001.F1.BR.01.01BR087
1993.F1.BR.93.93BR020
1989.F1.BR.89.BZ126
1989.F1.BR.89.BZ163
1993.F1.BE.93.VI850
1996.F1.FR.96.MP411
2002.F2.CM.02.02CM0016BBY
1997.F2.CM.97.CM53657
1995.F2.CM.95.MP255
1995.F2.CM.95.MP257
1997.K.CD.97.EQTB11C
1996.K.CM.96.MP535
2001.C.AR.01.ARG4006
1992.C.BR.92.BR025
1996.C.BR.98.98BR004
2004.C.BR.04.04BR013
1999.C.IL.99.99ET14
1991.C.KE.91.KNH1268
2000.C.BW.00.00BW07621
2000.C.BW.00.00BW16162
1996.C.ZM.96.96ZM751
2001.C.TZ.01.A125
1998.C.ZA.98.98ZA445
2001.C.TZ.01.A301
2002.C.ZM.02.02ZM114
1998.C.BW.98.98BWMO1410
1998.C.CN.98.YNRL9840
1993.C.IN.93.93IN101
1994.C.IN.94.94IN11246
2003.C.GE.03.03GEMZ033
2000.C.KE.00.KER2010
2002.C.TZ.02.CO3056
2004.C.ZA.04.04ZAPS161B1
2000.C.ZA.00.1119MB
2001.C.ZA.01.2134MB
2005.C.ZA.05.05ZASK243B1
1999.C.ZA.99.DU422
2005.C.ZA.05.05ZASK246B1
1997.C.TZ.97.97TZ05
2002.C.ZA.02.02ZAPS008MB1
1998.C.ZA.98.98ZA528
1996.C.ZM.96.96ZM651
1999.C.BW.99.99BWMC168
1998.C.TZ.98.98TZ017
1993.H.BE.93.VI991
1993.H.BE.93.VI997
1990.H.CF.90.056
1993.J.SE.93.SE7887
1994.J.SE.94.SE7022
1995.U.NL.95.H10986
1983.U.CD.83.83CD003
1990.U.CD.90.90CD121E12
0.07

Fig. S2


1998.C.TZ.98.98TZ017
1991.C.KE.91.KNH1268
1999.C.IL.99.99ET14
1992.C.BR.92.BR025
1998.C.BR.98.98BR004
2001.C.AR.01.ARG4006
2004.C.BR.04.04BR013
1999.C.BW.99.99BWMC168
2001.C.TZ.01.A301
2001.C.TZ.01.A125
1998.C.ZA.98.98ZA445
2000.C.BW.00.00BW07621
2000.C.BW.00.00BW16162
1996.C.ZM.96.96ZM751
2002.C.ZM.02.02ZM114
2002.C.ZA.02.02ZAPS008MB1
1997.C.TZ.97.97TZ05
1996.C.ZM.96.96ZM651
1998.C.CN.98.YNRL9840
1993.C.IN.93.93IN101
1994.C.IN.94.94IN11246
1998.C.ZA.98.98ZA528
1998.C.BW.98.98BWMO1410
2003.C.GE.03.03GEMZ033
2001.C.ZA.01.2134MB
2005.C.ZA.05.05ZASK246B1
2000.C.ZA.00.1119MB
2004.C.ZA.04.04ZAPS161B1
2002.C.TZ.02.CO3056
2005.C.ZA.05.05ZASK243B1
1999.C.ZA.99.DU422
2000.C.KE.00.KER2010
1985.D.CD.85.Z2Z6
1981.BE81
1984.D.CD.84.84ZR085
1999.D.TD.99.MN011
2001.D.CM.01.01CM0175BA
2001.D.CM.01.01CM0009BBY
2004.D.KR.04.04KBH8
2001.D.TZ.01.A280
1993.D.KE.93.MB2059
1997.D.KE.97.ML415
1999.D.UG.99.99UGB21875
2001.D.KE.01.NKU3006
1998.D.UG.98.98UG57128
1998.D.UG.98.98UG57131
1959.M.CD.59.ZR59
1990.B.US.90.US4
1983.B.US.83.RF
1989.B.EC.89.EC102
1983.B.GB.83.CAM1
2000.B.NL.00.671
2001B.UA.01.01UAKV252
1997.CAN97
2001.B.CO.01.PCM013
2001.B.CO.01.PCM074
2002.B.AR.02.02AR114146
1999.B.BR.99.BREPM11931
2004.B.JP.04.DR6089
1996.B.NL.96.H434
2005.B.JP.05.DR7060
2004.B.RU.04.04RU139089
1983.B.US.83.5018
2004.B.KR.04.04KMH5
2004.B.KR.04.04LHS6
1997.B.KR.97.WK
1986.B.NL.86.3202A21
1996.B.US.96.1299
1987.B.AU.87.MBC925
1990.B.US.90.US2
1990.B.TH.90.BK132.AY173951
1984.B.US.84.5019
1987.B.US.87.5113
1991.B.US.91.DH12
1998.B.US.98.98USHVTN3605
1996.B.US.96.5155
2004.B.AR.04.04AR151263
1987.B.US.87.BC
1997.B.US.97.1001
1990.B.US.90.US1
1986.B.AU.86.MBC200
1996.B.AU.96.MBCC98
2005.B.KR.05.05CSR3
1997.B.US.97.1013
1998.B.US.98.98USHVTN1925
2002.B.YE.02.02YE507
1986.B.GB.86.GB8
1998.B.US.98.98USHVTN8229
1998.B.TH.96.96THNP1538
1999.B.MM.99.mSTD101
1986.B.US.86.AD87
2001.B.UY.01.01UYTRA1092
1986.B.DE.86.HAN
1986.B.DE.86.D31
1988.B.GA.88.OYI
1983.B.US.83.5157
1989.B.EC.89.EC003
1988.B.US.88.WR27
1993.F1.BE.93.VI850
2002.F1.AR.02.ARE933
1989.F1.BR.89.BZ163
1989.F1.BR.89.BZ126
1993.F1.BR.93.93BR020
2001.F1.BR.01.01BR087
1996.F1.FR.96.MP411
1995.F2.CM.95.MP255
1997.F2.CM.97.CM53657
2002.F2.CM.02.02CM0016BBY
1995.F2.CM.95.MP257
1995.U.NL.95.H10986
1996.K.CM.96.MP535
1997.K.CD.97.EQTB11C
1983.U.CD.83.83CD003
1990.U.CD.90.90CD121E12
1993.H.BE.93.VI991
1990.H.CF.90.056
1993.H.BE.93.VI997
1993.J.SE.93.SE7887
1994.J.SE.94.SE7022
2001.A.SN.01.DDJ369
2001.A.SN.01.DDI579
1996.A.SN.96.DDJ360
2004.A.ZA.04.SK162B1
1992.A1.UG.92.UG029
1994.A1.SE.94.SE7253
1998.A1.UG.98.98UG57134
1999.A1.KE.99.KNH1088
1995.A1.SE.95.UGSE8131
1992.A1.UG.92.92UG037
1999.A1.UG.99.99UGG03379
2000.A1.KE.00.NKU3005
2001.A1.TZ.01.A195
1999.A1.KE.99.KNH1135
1995.A1.SE.95.SE8538
1997.A1.KE.97.ML752
2001.A1.TZ.01.A173
1997.A1.KE.97.ML605
1997.A1.TZ.97.97TZ03
1998.A1.UG.98.98UG57135
1995.A1.SE.95.SE8891
1999.A1.UG.99.99UGA07072
1998.A1.UG.98.98UG57136
2000.A1.KE.00.MSA4079
1986.A1.KE.86.ML170
2000.A1.KE.00.KER2008
1998.A1.UG.98.98UG57142
1992.A1.RW.92.92RW008
1994.A1.SE.94.SE7535
2001.A1.TZ.01.A341
1997.A1.TZ.97.97TZ02
1994.A1.KE.94.C23
1993.A1.RW.93.93RW024
1992.A1.RW.92.92RW025A
1993.A1.RW.93.93RW037A
1997.A.CD.97.KTB13
2002.A.CD.02.KTB035
1997.A.CD.97.KCC2
1960.DRC60
1994.A2.CY.94.94CY017
1997.A2.CD.97.97CDKTB48
1985.BE85
10.0 years

Fig. S3


effective population size
lower 95% HPD on M group TMRCA
1.0E6
1.0E5
1.0E4
1.0E3
1.0E2
1.0E1
1.0E0
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
year